Proceedings of the 56th Hawaii International Conference on System Sciences | 2023

Big Data guided Resources Businesses – Leveraging Location Analytics and Managing Geospatial-temporal Knowledge

Shastri L Nimmagadda
Curtin University, Australia
shastri.nimmagadda@curtin.edu.au

Andrew Ochan
Petroleum Authority of Uganda
andrew.ochan@pau.go.ug

Torsten Reiners
Curtin University, Australia
t.reiners@curtin.edu.au

Neel Mani
CAIR, D. S. Vishwavidyalaya, India
neelmanidas@gmail.com

Abstract

Location data rapidly grow with fast-changing logistics and business rules. Due to fast-growing business ventures and their diverse operations locally and globally, location-based information systems are in demand in resource industries. Data sources in these industries are spatial-temporal, with petabytes in size. Managing volumes and various data in periodic and geographic dimensions using the existing modelling methods is challenging. The current relational database models have implementation challenges, including the interpretation of data views. Multidimensional models are articulated to integrate resource databases with spatial-temporal attribute dimensions. Location and periodic attribute dimensions are incorporated into various schemas to minimise ambiguity during database operations, ensuring resource data's uniqueness and monotonic characteristics. We develop an integrated framework compatible with the multidimensional repository and implement its metadata in resource industries. The resources' metadata with spatial-temporal attributes enables business research analysts a scope for data views' interpretation in new geospatial knowledge domains for financial decision support.

Keywords: Resources Industry, Heterogeneity and Multidimensionality, Data Warehousing and Mining, Big Data.

1. Introduction

Resources data need interaction and analysis with location-based information systems [6, 9, 15]. The business research explores spatial-temporal resource data's heterogeneity and multidimensionality challenges and presents a business case through modelling and management. The existence of hundreds of years of volumes and varieties of data in these industries motivates us to undertake the current research [12, 18]. From exploration to marketing through production stages, industries generate enormous volume and variety of data with information flow among various operational units of exploration businesses [1, 12]. We characterize the resources' data as associated with oil and gas and mineral and mining entities, often periodic and geographic. The periodic dimension consists of composite attributes such as day, month, quarter or year in different hierarchies. Accessing and managing accurate information from geographically located operational units is challenging for making timely trade decisions. Considering the issues associated with existing tools and technologies of the resources' businesses [2, 5], IS based artefacts and location analytics are investigated. From data management and science perspectives, logical storage is needed to support the integration process, besides managing hardware and software platforms. These tools have the flexibility to accommodate multiple dimensions and their attribute instances in a variety of business entities [2, 4, 6]. The importance of location analytics research has implications regarding business intelligence, data warehousing, scalable data lakes, and operational analytics. We can fuse these topics in a single integrated framework. This framework can work in any application domain, whether business, healthcare, or environmental ecosystem. The present investigation presents Big Data characteristics in an Australian resources industry and how the heterogeneous and multidimensional data sources are articulated with spatial dimensions and their collaborations through mapping and modelling methodologies [1, 5]. The research objectives, the need and significance of Big Data in the resources industry with attainable goals by robust methods are described in various sections. Data mining, visualisation, and interpretation of resources metadata are discussed, focusing on spatial-temporal data analytics done in a repository environment.

2. Issues, Challenges and Motivation

The best practice of data warehouse development can increase the value of businesses and their stakeholders. The research lays the groundwork for a data warehouse or repository to grow and adapt to business needs and requirements. As discussed by [15], Design Science theory has motivated us to draw multidimensional constructs or conceptual models and logical models, based on which more dimensions are manageable, making it easy to view or compare the

URI: https://hdl.handle.net/10125/103247
978-0-9981331-6-4

fused data through data cubes. However, the approach discussed in [9] is still valid because the business data and information are stored and managed in a single location. Security of sensitive information and restricting its usage in relevant industries are valuable to managers. However, for accessing the data through queries, relational databases pose limitations. Managing complex SQL queries through joins of multidimensional tables, including handling the data views by visualisations and interpretation of Big Data, are other challenges. Depositing data in a single location and enhancing their security are advantages of relational databases. Grouping similar information, speedy access, easy maintenance and better performance are positive features of multidimensional construct designs and models. The integrated frameworks can cater multiple digital ecosystems and business applications. For example, logistics and supply chains and their affected digital ecosystems can be added to the frameworks.

Resources data are broadly spatial-temporal. Location attributes are referred to as spatial attribute dimensions. The accumulation and existence of spatial-temporal heterogeneous and multidimensional data sources in several resource business organizations have motivated us to carry out the current research work [16]. Dimensional data stored in the resources companies are in bits and pieces, which may not be compatible with the data integration process, besides constraining the access protocols on various software and hardware platforms [14, 16]. From a Big Data perspective, previous studies indicate that data warehousing and data mining technologies appear promising for developing effective data storage solutions in resources industries, especially when the industry is going through turbulent business situations [17, 2]. The early research on Big Data, data warehousing, and mining methods, which are implementable in various industry applications, have guided us to examine the existing tools. We propose to redesign the data constructs and models in resource business contexts. A classic example of the airline industry's data warehouse design and implementation is discussed in [9]. Underlying ontologies are described in a case study in the petroleum industries in [14]. Operationally, accessing and integrating the specific data and information required for exploration and field development activities are in high demand, incorporating the spatially varying drilling and production entities and dimensions. Earlier data structures in these business systems are obsolete, with poor data qualities [16, 17]. Besides, the data retrieval from volumes of data sources in such operational contexts is complicated [16]. Structuring business data is intricate in large size repositories, where periodic and geographic dimensions may need more joins to bring the query results into one table [14]. Data mining is a critical motivating task especially assessing multidimensional exploration and production data that exhibit schematic, syntactic and semantic heterogeneities. Adding the spatial-temporal dimensions in every resource-based data model has significance but needs attention in relating the structuring and integration process with other associated attribute models. The process can establish the connectivity between exploration and production entities of associated oil and gas-bearing sedimentary basins [12]. Spatial scientists and resource explorers may be interested in real-time data processing and monitoring through innovative logical and physical structures with new business rules. For example, it may be necessary to record or document the current *petroleum permit status* or its location history in a repository system. Business managers propose documenting the current prices of all petroleum products or exploration costs and their periodic fluctuations, including their spatial business variations [7]. A centralized repository can share processed data and information with working centres, individuals and operational teams functioning locally and globally, so multiple users can make timely decisions [10]. The approach has a broader scope, but the present study is limited to the design and development of data models, addressing their implementations in spatial-temporal dimensions, including roles in refining the data structuring process in the application framework.

3. Research Questions and Objectives

We focus on the performance of databases, easing data access, including using and reusing stored data structures in multiple domains. In this context, we take the guidance and advantage of the Big Data paradigm [1, 16] to design the research objectives:

1 Develop data constructs and models, distinguishing the instantaneous and historical data associated with volumes and various Big resources data. Analyse changes occurred in the business rules affecting Big Data modelling. The manner spatial-temporal characteristics affect the cardinality of relationships in Big Data is investigated.
2 Present the spatial-temporal Big Data in cubes. Analyse how Big Data characteristics influence data modelling and facilitate resolving the complexity of spatial modelling in resources' business contexts.

Based on the research purpose and objectives, the following research questions (RQ) are designed:

1 How do we structure the spatial-temporal resource data, considering their heterogeneity and multidimensionality?
2 How do we interconnect business information systems, implement them in upstream industries,

and explore metadata views to interpret in new knowledge domains for locating oil and gas prospects and managing them through Big Data?

Business activities and functions of resources industries are diverse in Australian resources industries. We design the current research objectives, keeping in view the Big Data, especially the petabyte size of data volumes involved with the spatial-temporal attributes and their instances. The research objectives can resolve the challenges of a competitive and distributed business environment locally and globally.

4. Significance of Location Big Data

Australia's economic growth depends on exploring and exploiting Western Australia's (WA) natural resources [12, 7, 5, 17]. We focus on modelling the location-based data sources of WA's resources businesses to assess the viability of exploration and use of new technology tools in the resources industries. They are typically spatial-temporal, and their spatial analytics is crucial for geographically establishing the growth of exploration and production [3, 12]. The information needed for current and future resources' predictions, including the speed at which the warehouse repository can deliver user query results, are different tasks of the present research. For example, geologists can display *formation tops* (geological attributes) for immediate use by reservoir engineers, calibrating their models and ensuring that spatial-temporal information is current with updates of exploration and production entities. Managers can investigate drillable-well locations, exploring at different spatial dimensions.

The current models and insights of previous researchers' perspectives can be incorporable into the proposed integrated framework. By adding growing insights of location analytics into the integration process, we can increase robustness of the framework by fusing new exploration and production entities. Location-based seismic and other geophysical attribute instances have a bearing on the current mapping and modelling approach. We have acquired volumes and varieties of geophysical data as a part of exploratory drilling campaigns in the study areas. Each volume consists of several gigabytes of data with different G & G contexts and their variabilities. Seismic is one of the G & G methods of prospecting and exploration of oil and gas deposits [2, 12]. Exploration staff optimize their analysis by knowing which seismic profiles in the field have the most appropriate acquisition and processing abilities to deliver quality interpretation results. They use these results to plan for additional surveys when they are found inappropriate to use and interpret for prospect analysis [7]. Geological and geophysical (G & G) exploration scientists may need to access the drilled-well data acquired in different locations and periods to facilitate the extraction of new geological knowledge relevant to production histories and other geophysical entities. In the entire energy production and marketing cycle, Exploration and Production (E&P) is a primary phase of the Oil and Gas business that involves searching for and extracting oil and gas deposits. In other words, the exploration task involves deploying exploration and prospecting methods to help locate promising sweet spots for oil and gas drilling and extraction. The drilling staff may interpret the exploration data to cognize the geological formation qualities and their depths in a total field area. We analyze forecasting of similar production profiles from historical exploration data, where active location permits exist. As per RQ 1, the authors examine the data issues and resolve the concerns using mapping and modelling methods. Exploration data acquired in different field seasons and locations are shown in Figure 1, three exploration areas (I, II, and III) corroborating the attributes of several surveys, the number of wells drilled, and many oil and gas wells. Each number describes several surveys, drilled wells and permit licenses to do exploration work. Bubble plots of yearly-accumulated data volumes are referred to number of surveys, drilled wells and permit attributes acquired by companies. Big size of bubble indicates magnitude of the attribute instance (Figure 1).

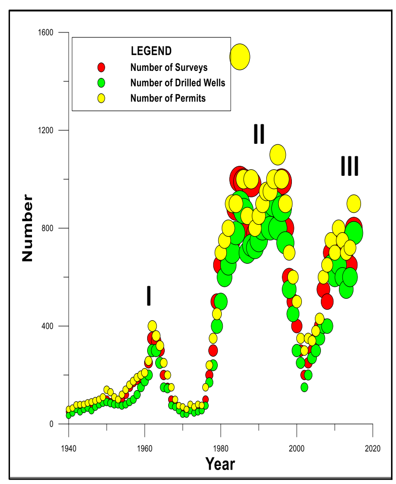

Figure 1: Exploration business data acquisition

Data quality is the critical measure of data condition, such as accuracy, completeness, consistency, and reliability, ensuring the dataset is current in the prospect analysis. In the present research, we interpret data sparsity, heterogeneity, multidimensionality and granularity in bubble plots from Big Data sources (as envisaged in Figure 2). In a 2D bubble plot, each bubble's diameter varies in size, providing a way to represent additional dimensions in the data. The Big

Data characteristics are evaluated in periodic and geographic dimensions by interpreting increasing bubble size, with periodic growth of attribute strengths and change of data characteristics.

For acquiring the location-based geophysical data, we lay out different profiles or grids that are significant in the investigating areas. Despite the data qualities and characteristic properties that deliver quality exploratory drilling operations and prospective locales, Big Data is a leading-edge technology in précising exploratory drilling locations [2, 12, 16, 17].

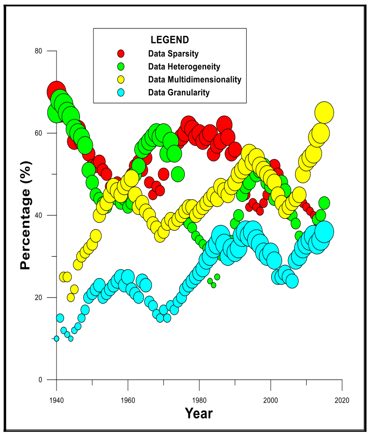

Figure 2: Data quality representing characteristics

Similarly, the bubble plot views are schematised in Figures 2 and 3 with yearly-accumulated Big Data volumes that referred to number of surveys, drilled wells and permit attributes in Figure 1.

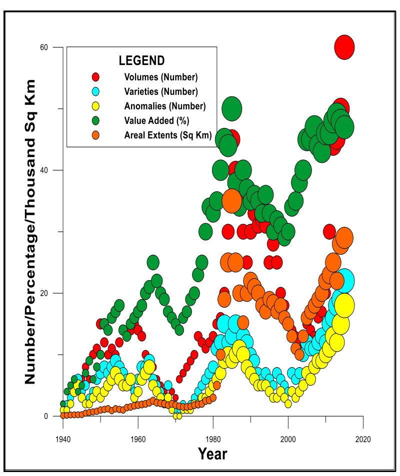

Figure 3: Periodically varying data characteristics

We describe the Big Data characteristics in multidimensional logical star-schemas to connect and integrate with related logical schemas of E & P business metadata [16]. Areal extents of Big Data, value addition and anomalies construed from Big Data volumes and varieties are graphically presented in Figure 3. The queries generated from these repositories can facilitate present and exchange the information among regional, project managers and unit-level operations of resource businesses. Models generated using statistical correlation, regression analysis, and other mining schemes can go with predictive models that can provide knowledge of future forecasts of resources [11, 16]. As described in Figure 4, a knowledge base system is designed to combine multiple roles and activities of exploration, drilling, production and technical entities of the oil and gas business through ER constructs with several rules imposed on them [12, 14, 15]. Similarly, new perceptions of exploration and production data are updated in the framework with new investigating areas.

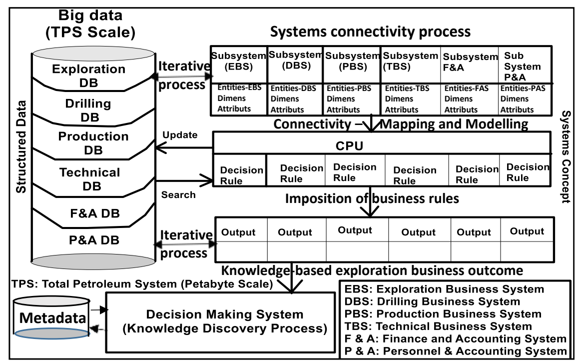

Figure 4: Systems connectivity – repository-building process

The repository is a data library for data mining for reporting and analysis (Figure 4). For storing E & P data and information, the storage is a vital site that can centralise information with ease of access by programmers, coding professionals, testing, and debugging support staff. The operational database system covers the present-day activities significantly and functions of exploration, drilling and production entities. For real-time business data manipulations, petroleum information system analysts engross the Big Data for better managerial decisions and resolving complex company queries construed by outside parties [16, 17]. Initially, a conceptual model is described before envisaging a logical model for connectivity among exploration, drilling and production entities and their attributes in the existing resource business databases (Figure 4).

The resource data grow with spatial dimensions into Big Data. Big Data moves faster among various hardware and software modules in an integrated project environment based on the size of data volumes and varieties of resource businesses [5]. We monitor the accuracy and validity of data in these modules such that the legality of systems and their connectivity are well understood based on location attribute dimensions. The connectivity, as demonstrated in Figure 4 among various entities, is further explored, integrating with other dimensional models that represent the spatial-temporal dimensions. For example, regarding the cost of exploration and discoveries made at various periods in different geographic regions of Australia and basin settings, plot views can investigate patterns and trends in the data through visualisation and interpretation [7].

5. Data Modelling Methodologies

Regarding theoretical foundations, the current research drives the constructs and models based on theoretical design science approaches. We identify key concepts to evaluate and explain relevant systems as part of theoretical foundations. The way the design science theory best fits location-based business contexts and their constructs and models, is enlightened. As per RQ 1, various modelling methods are analysed. Volumes of exploration and production are assets of the resources business organisation. Prospect analysis must document and maintain assets geographically and periodically, ensuring knowledge is current, even by integrating the old datasets with the existing ones.

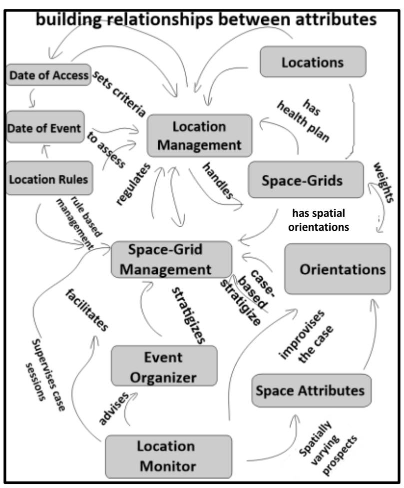

Figure 5: Building spatially varying ontologies

As conceptualised in Figure 5, we have built spatial ontologies to support and establish the Design Science Research (DSR) theory and associated concepts and contexts of spatial entities and dimensions, including objects. The associations among interconnected space-grids, grid orientations, grid densities, including spatial attributes are also considered. Location management, spatial event organisers, and location monitors (GPS) are interconnectable through ontology descriptions to exploration entities. These entities have spatial or location bearing that makes up the geospatial relationships in Petroleum System domains.

5.1 Storing the Resources Big Data

For location modelling and management, schematic and semantic-based tools are usable, collaborating information, including implementable and interoperable navigational data structures at multiple locations. Ontology guided spatial information system uses several data relationships across various domains, and their integration can facilitate collaborating knowledge-based decision-making process. We have initially built conceptual models and entity-relationship (ER) logical models that can connect different entities associated with business activities and functions of oil and gas entities [14]. It is a graphical representation of an entity-relationship diagram and a detailed logical depiction of various entities in an exploration business scenario, as presented in Figure 6. Similarly, multiple data entities and attributes identified for petroleum exploration and production are conceptually represented in graphical forms to understand cardinalities explicitly [14, 16, 17].

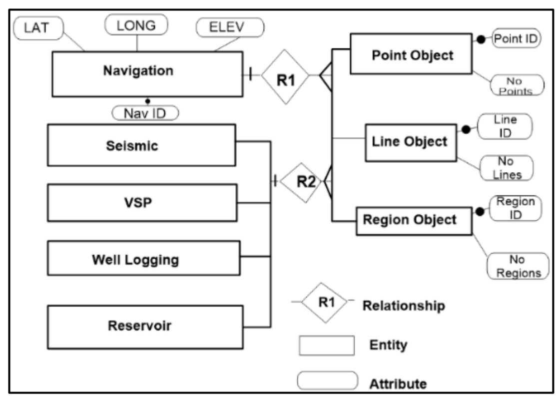

Figure 6: Conceptual modelling of exploration business with navigation (location) dimension

From the information system design perspective, conceptual models describe the pathways of knowledge-based interaction between multiple entities, dimensions and their attributes (Figure 6). Several objects, such as point, line (profile) and region, are interpretable and connectable to several core entities of the exploration dealings. The business objectives could be functional and purposeful, interpreting various contextualized entities of the exploration businesses. In dimensional modelling, navigation, seismic, and Vertical Seismic Profiling (VSP), including well logging and reservoir entities and their associated attributes, are applicable. For building fine-grained multi-dimensional logical data structures, ER models are connectable with attribute dimensions and their instances of associated domains of exploration businesses [14, 17]. The next stage is building a repository with data acquisition, data structuring, storage and data mining components [4, 16]. Visualisation and interpretation are other value-added artefacts of the integrated knowledge management system. Sequences of events are planned, firstly, the data acquisition from several data sources and then importing and or exporting data into repository systems [5, 16]. Secondly, the design of relational, hierarchical and multidimensional data structures is meant to accommodate them in the large-size storage system. Thirdly, several business rules are designed to integrate with various conceptual and logical data structures. We illustrate two simple examples that demonstrate the spatial-temporal aspects of the current research problem. An upstream oil and gas business entity requires documenting and storing periodic and geographic attribute instances with deliverable surveys [12, 16]. For example, the business rules are: (1) Each exploration holds at least one production license; (2) Each exploration license may hold a navigational entity. The upstream business desires to store petroleum permit information, including licenses. The business rules are:

1. The relationship changes one–to–many to into many-to-many since the exploration over time varies significantly. An associative entity must hold the dates when location records need connections to survey profiles (Figure 6).
2. The data relationships can change in spatial-temporal dimensions. One contractor holds one license, with many surveys and operations. In the exploration/survey example, the association is said to be transferable. A survey held by an exploration entity can be transferred to another type of exploration to integrate with the space domain. In other words, when a prospect is explored by one variety of exploration techniques, the prospect can now be validated by another variety of exploration methods.
3. Survey permits belong to one and only contractor. A new permit order is placed if the contractor with existing licenses wants to add other nearby permits. A simple criterion is to regulate when deciding if one-to-one or one-to-many relationships develop into many-to-many relationships in spatial dimensions.

When referring to the conceptual model in Figure 7, the question may arise: how can one distinguish between the current and past survey data sources in spatial-temporal dimensions in unstructured Big Data situations? The detailed conceptual model does not separate their survey records. A conceptual model is designed to understand how the navigation, geophysical profiles, and linked line, points and regions are connected through shot point locations (Figure 7). Various location-survey attributes have affected the data relationships modelling and its structuring cardinalities.

Navigational attribute dimensions are connectable to the multiple areas through survey lines and shot point locations (where G & G data are acquired in the field).

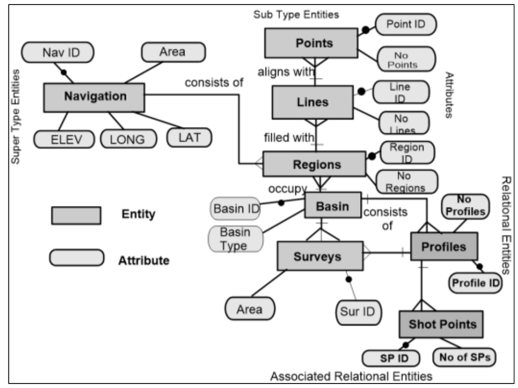

Figure 7: A conceptual model for location-survey problem in an exploration entity

The same contractor may have several leases or permits in the same field. The contractor may possess multiple fields or basins with detached location attributes. We incorporate updated business rule attributes in the modelling process.

5.2. Using a dependent entity for location-based exploration business data

Several dependent entities are created to simplify data models in many-to-many relationship situations. For example, many surveys have many profiles and many shot points, but each point describes the shot location. Each unique survey profile can be in a survey of the region under investigation. Business rules play key roles in modelling the data relationships among spatial-temporal data. Surveys depend on the navigational entity's existence, and thereby, the existence is dependent if it has a mandatory foreign key from the location dimension. In this case, we interpret one or more business rules: One-to-one relationships become one-to-many or many-to-many. The data to be stored in the repositories include the instances associated with the exploration locations, exploration details, and petroleum production data, including their periodic attributes. More precisely, varieties of G & G data are acquired, such as well-data from drill sites, well-site installations and drilled-well expenditure in different locations and periods. The conceptual models drawn with an exploration entity or dimension describe attribute keys (Figures 6 and 7), such as exploration ID, Navg ID, details of surveys, exploration name and start date of exploration. This model cannot deliver the required data views for interpretation without incorporating relevant location data attributes; the present model must be adaptable to meet the new requirements. The changes are shown in Figures 6-7. The new entity or dimension, described from exploration histories, has a dependent entity that keeps track of different navigations and surveys over time. One needs to make decisions or choices at this stage about the recording of survey location, the number of surveys, exploration costs, production rates, and drilled-well pressure attributes. There may be two possibilities: (1) Spatial-temporal changes in exploratory drilling or (2) Changes in geological entities or dimensions based on geography or location dimensions. The choice depends on the frequency of change and the importance of recording every change occurring in the data relationships, irrespective of any attribute dimensions.

5.3. Changes in data relationship cardinality

In the data modelling process, one-to-one and one-to-many relationships can turn into many-to-many data relationships based on conceptualisation and contextualisation features [12, 14]. The contextualisation aspect describes the location dimension hierarchy. It is always not right rules imposed on business situations, in which cardinality plays a key role in the modelling process [4]. Two reasons can cause the cardinality change (1) change of business rules (2) repositories ensure constant updates based on location hierarchies.

5.4. Changes in business rules

Business rules meant to govern businesses are captured along with their relationships with the systems or organisations' entities and/or dimensions [17]. Resources data in upstream businesses swiftly grow as per business rules. The changes in business rules often affect the data models. For example, one or more contractors hold surveys or licenses associated with data acquisitions. In this context, the company stores a list of survey business rules and the location data details that depict data acquisition, including prospect analysis of the upstream exploration industry. Awarding licenses or permits to different contractors is an ongoing business process. Two weeks after successfully implementing the tables in the relational databases, another contractor, with the approval of the management of the resources company, decided to acquire more surveys in the nearby investigation areas (composite spatial dimension) and add them to the existing databases for building additional exploration knowledge. Such business situations are unpredictable and unavoidable in the resources businesses, implying that the relationships grow based on spatial-temporal dimensions [16, 8]. Additional data attributes need to be stored, such as contexts when a contractor acquires a survey, suggesting that complex data relationships are resolved with associative entities.

Changing business rules by many contractors or employees associated with employers for that survey is a simple change to the problem. Still, it results in the review of the entire database structure. We investigate the business rules periodically and cognitively in space dimensions to assess the probability of exploration changes in the short or medium term.

5.5. Modelling Multiple Spatial-Temporal Dimensions

In addition to entities and dimensions, we can use objects for aligning the operational database designs. We convert the conceptual ER models into various logical dimensional and object-oriented schemas that narrate the structural description of different objects created by the user, such as tables, queries, forms, reports, views and constraints [14]. For constructing databases, ontology modelling is still a modern approach through ER representations of entities and their relationships. However, for compatibility and flexibility, dimensional schemas are updated to generate a warehouse repository that accommodates current multidimensional data. Scalability and efficient use of storage and analytical processing are added advantages.

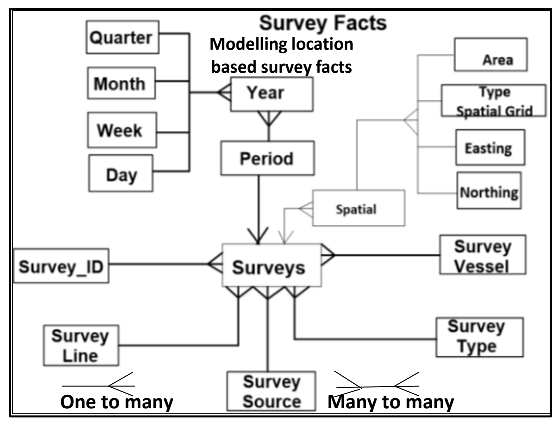

Figure 8: Multidimensional star schema, articulating survey facts, involving spatial-temporal attributes

In addition to dimensional models, we need an organized data structure to optimize and use business intelligence through numerical factual data relating to the surveys separate from descriptive or dimensional data [16]. One star schema uses a single large fact table to store the survey data with one or more dimensional tables surrounding the factual survey data. The location and period are dimensional tables surrounding these survey facts tables (Figure 8). The star schema interprets the structure of the dimensions of resources' data. The most common database models that define the data relations have one-to-many or many-to-many relationships. Cross-reference key attributes to connect the tables that represent the relationships between entities. Primary and foreign keys provide easy access to the databases. Database contents vary with both location and time attribute dimensions [10]. We describe a multidimensional diagram involving the spatial dimensions drawn for an upstream business (Figure 8). Further, for accommodating complex G & G domains and petroleum systems, fact constellation schemas are considered for articulating data tables logically [14, 15]. The fact constellation schemas can be adaptable or functional in multiple domains of the resources business information systems for building repository systems.

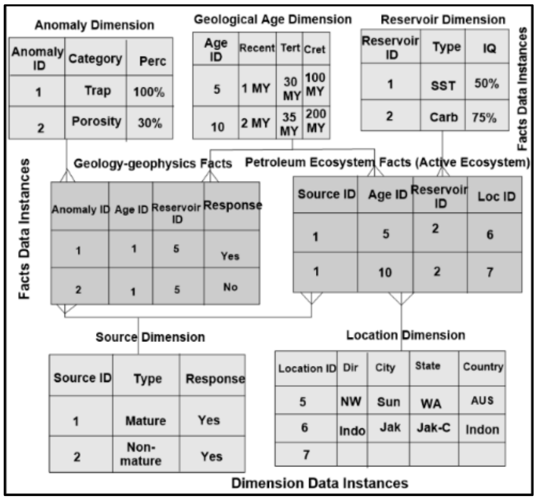

Figure 9: A constellation schema connecting G & G and petroleum ecosystem domains

The schematic view of constellation schema depicts a compilation of multiple fact tables with common dimension tables, implying that location and periodic dimensions are closely relatable to G & G domains, including elements and processes of the petroleum information systems (Figure 9).

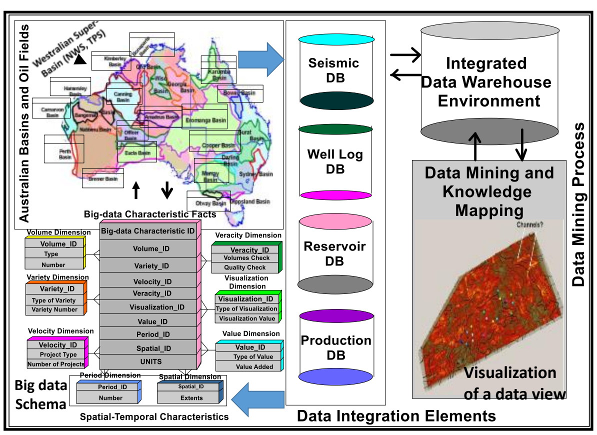

Figure 10: Implementation workflow

The resources companies doing business at several geographic locations in Australia and overseas do transactions in different periodic instances. The periodic dimension is made relevant to both location-based exploration data and oil and gas discovery data facts, implying that the period dimension is linked to different operations and activities of the resources industry. As shown in an implementation workflow in Figure 10, Big Data of Westralian Super Basin has volumes and various attribute dimensions and facts, enabling us to apply and place all the data in a single repository. The process allows data mining and visualization done at different locations to make swift financial decisions [12].

Several resources' data tables are documented with location attributes and their instances. The common attribute dimensions for connecting all the fact tables in each volume are (1) Number of surveys (2) Number of drilled wells (3) Number of Permits (4) Exploration Type (5) Number of Basins (5) Number of Companies and Contractors. Some of them are presented in Figures 1-3. The data structures involving the location explore these volumes for connections among multiple attributes of the resources data warehouse. The data volumes acquired on resource businesses and their documented data instances need continuous periodic updates revising the data schemas. The data relationships are denormalized for fine-grain schemas to yield finer data views at par with user queries. As per RQ 2, data volumes and their cubes are implemented. One way to view fine-grain data models is through a multidimensional cube (Figure 11). Each cell contains one or more attributes, or in dimensional modelling, attributes are categorized from the raw data [13, 16]. Attribute dimensions such as locations, periods, surveys conducted, contractors, and wells-drilled are represented in aggregated data views (Figure 11).

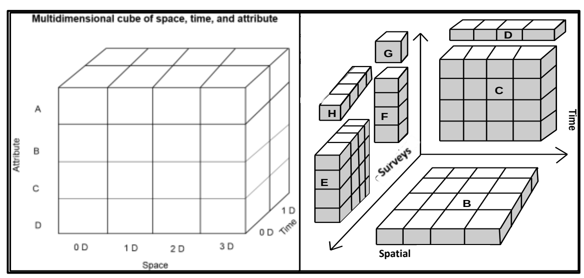

Figure 11: Data views from Big Data cubes showing slices as aggregate data views B – G

In a dimensional cuboid structure, each cell holds data relevant to the intersection of all its dimension values. For example, a cell might contain several drilled wells in a particular location, period and basin with a specific number of oil and gas producing horizons [12]. Such spatial-temporal data are typically aggregated (Figure 11) with appropriate time intervals, yielding a large volume of equally spaced time-series data in different investigating areas. Such data are analysed using mining schemes developed in modern time-series analysis and other statistical tools [11]. Several aggregates ((B) - (H)), as shown in Figure 11, facilitate the users' perception of visualisation and interpretation. Visualisation of metadata cubes and their views add values to the existing interpretation, replacing it with new knowledge of exploration and production data. If all survey-drilled-wells-permits are fused, we can arrive at cumulative ten volumes in the framework (Figures 10 and 11), with a total Cube Metadata Size that could have been 10x1100x350x300x360 bytes. After adjustment in complete transactions, the total database size could be 415 GB if considered in a particular period dimension. We dig the schedules, and each day drilling a borehole if completed ahead of warehouse construction. In addition, we document the operations of each truck or rig used in the drilling carried out at different locations. Each lease is given a unique permit number. For efficient operations, the movement of drilling rigs is documented between adjoining concessions and proposed borehole plans [7]. Access to the databases is requested from corporate metadata to create data views from working groups of exploration, drilling, production and marketing entities in remote sites.

Location-based resources data analytics: We build location-based queries and data views from repositories. Certain drilled wells and or seismic profiles are chosen that are currently focused on interpreting seismic data for exploring geological structures, isopachs and reservoir engineering models [2, 12]. We have presented the map views in Figures 12-13 to interpret oil and gas prospects. We present data views from metadata cubes for new knowledge interpretation through visualisation. Certain locations and periods with high production rates are interpreted as having a specific geographic bearing, where seismic and drill-well campaigns are active in such locations and periods. The seismic, drilled-well and permit license information are needed for a particular location to accommodate the plot and high-quality map views, as described in multiple graphical windows [8]. In this context, the use and reuse of the data structures are evaluable through systems' approach and the knowledge of connectivity that led us to interpret unexplored resources in location attributes [16, 17]. We have computed G & G data structures, which are relevant to exploratory drilling campaigns and compute metadata usable for data mining and visualisation (Figure 12). We computed volumes or cubes of attributes to present their structural variations based on location attributes. Structural mapping is identifying and characterising geological structural footprint or expression and exploring oil and gas traps. These are basic attribute maps required by resources analysts, explorers *and* production managers to manage oil and gas prospects. Geological structure, formation thickness (isopachs) and reservoir thickness maps are presented with location-based coordinates "northing" and "easting" attributes. Drilled wells are posted on the flanks of the geological structure, as shown in Figures 12 a and 12 b. Drilled well locations are plotted in the geological structure map visualisation, in addition to northerly direction (N) and color scale attributes. Location-based map view implies attribute mapping and modelling, done with northing and easting coordinate information. In addition to structure attribute maps, we need isopach maps that visualise stratigraphic thicknesses between upper and lower horizons that reflect the thickness of deposited bed in the study area. We have built location-based geological formation thickness map views (Figure 12b), ascertaining the structure and drainage pattern attributes.

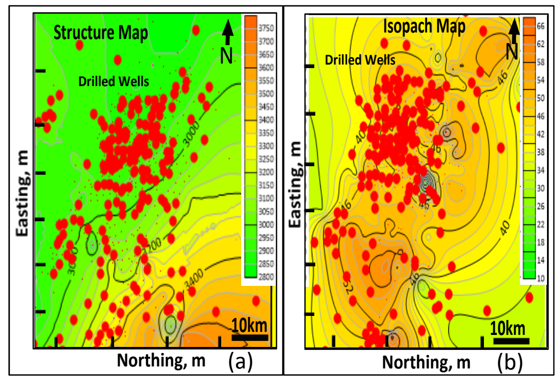

Figure 12: (a) Location-based geological data structure (b) Location-based formation thickness

These maps are often used by drilling campaigners and oil and gas explorers to make financial assessments and investments. As per RQ 2, G & G metadata are analysed to assess the models and location-based map views and how the geological formation thicknesses can be presentable to discern the thickness variations, the direction of sedimentary inputs, and drainage patterns in the investigation areas [2].

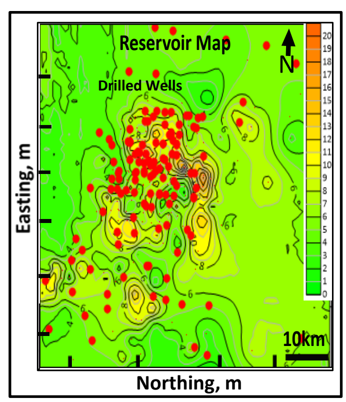

Figure 13: Location based reservoir modelling

RQ1 and RQ2 are addressed using geospatial mapping and modelling, including implementation methods. We present the location-based reservoir thicknesses and potential areas of prospective locales for reservoir modelling and field development (Figure 13). The bubbles located in the map views indicate drilled wells with proven oil and gas fields. Interpretation of map views indicates the promising existence of oil and gas reservoirs and estimates the extents of reserves in different thicknesses in a proven field (Figure 13).

6. Conclusions and Recommendations

The oil and gas business data comprise composite dimensions, such as geographic location and *period* features, with detailed periodic attributes and their instances. Both spatial-temporal data structures are complex and require more joins to bring the query results into one table. We demonstrate the use of these attributes in the Big resources Data modelling and their integration in a multidimensional dimensional repository environment. The performance and ease of access of repositories, as per user queries, are assessable using mapping and modelling methodologies. Big Data analysis provides interesting knowledge-based location data trends and patterns from multi-dimensional metadata cubes. These trends are represented in various knowledge-based map visualisations for oil and gas prospect interpretations. The spatial-temporal attributes are advantageous in the warehouse structuring of the resources data, despite geographic dimensions do affect the cardinality of the data relationships. Business rules described in resources' data structures are flexible enough to ensure the ease of use and reuse of structures. In the current application domain research, Big Data demonstrates the impacts of the resources industry, keeping in view the multidimensionality and granularity of the exploration data. The movement and integrity of data are contented with multiple resources' projects. Spatial dimensions of Big Data have a significant impact on visualization and interpretation of new knowledge that adds value to the existing resource projects. Prospect analysis is successful using location-based models and their map views. Recommendations include building a case study based on research findings and expanding theory and workflows. The IS constructs and models and their scopes are extendable in associated domain applications for new opportunities of location analytics in many other basins.

References

[1] Gandomi, A. and Murtaza Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics, International Journal of Information Management, Volume 35, Issue 2, Pages 137-144, https://doi.org/10.1016/j.ijinfomgt.2014.10.007.

[2] Castañeda Gonzalez, O.J., Nimmagadda, S.L, Cardona Mora, A.P, Lobo, A, and Darke, K. (2012). On Integrated Quantitative Interpretative Workflows for interpreting structural and combinational traps for risk minimizing the exploratory and field development plans, *Bolivarian Geophys Symp* proceedings, Colombia.

[3] Cleary, L, Freed, B, and Elke, P. (2012). "Big Data Analytics Guide: 2012", Published by SAP, CA 94607, USA.

[4] Salem, A., Ben-Abdallah, H. (2015). The design of valid multidimensional star schemas assisted by repair solutions. *Viet. J Comput Sci* **2,** 169–179 (2015). https://doi.org/10.1007/s40595-015-0041-1.

[5] Debortoli, S. Muller, O. and Brocke, J.V. (2014). Comparing Business Intelligence and Big Data Skills, BISE – RESEARCH PAPER, DOI 10.1007/s12599-014-0344-2, Springer Fachmedien Wiesbaden, 2014.

[6] Martyn, T. (2004). Reconsidering Multi-Dimensional schemas. SIGMOD Record. 33. 83-88. 10.1145/974121.974136.

[7] Dodds, K. and Fletcher, A. (2004). Interval probability process mapping as a tool for drilling decisions analysis – the R&D perspectives, *The Leading Edge*, Vol. 23(6) (pp. 558-564).

[8] Dunham, H.M. (2003). Data Mining, Introductory and Advanced Topics, Prent. Hall Public, 10-200p.

[9] Gornik, D. 2000. Data Modelling for Data Warehouses, Rational Software White Paper, www.rational.com/worldwide.

[10] Gregersen, H. and Jensen, C.S. (2002). Conceptual Modelling of Time-Varying Information, http://powerdb.net/database.

[11] Gupta, C.B and Gupta, V. (2009). An introduction to statistical methods, 23rd Revised Edition, Vikas Publishing House, New Delhi.

[12] Longley, I.M., Bradshaw, M.T., and Hebberger, J. (2001). *Australian petroleum provinces of the 21st century*. In M.W. Downey, J.C. Threet , and W.A. Morgan (Eds.), Petroleum provinces of the 21st century, AAPG Memoir, 74, 287-317, USA.

[13] Meersman, R. A. (2004). Foundations, implementations and applications of web semantics, parts 1, 2, 3, lectures at School of Information Systems, CBS, Curtin Uni, Australia.

[14] Moody, L. D and Kortink, M.A.R. (2003). From ER Models to Dimensional Models: Bridging the gap between OLTP and OLAP Design, Part1 and Part 2, *Journal of Business Intelligence*, Summer Fall editions, Vol. 8(3), http://www.tdwi.org.

[15] Peffers, K., Tuunanen, T. M. Rothenberger, M. A. and Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research, Journal of Mgt. Info. Sys, 24:3, 45-77, DOI: 10.2753/MIS0742-1222240302

[16] Rudra, A. and Nimmagadda, S. (2005). Roles of Multidimensionality and Granularity in Warehousing Australian Resources Data.. 10.1109/HICSS.2005.527.

[17] Ronald G. R. (2011). "Know-How Models: How Business Rules, Decisions, and Events Relate in True-to-Life Business Models" *Bus Rules Journal* Vol. 12 (11), http://www.brcommunity.com/a2011/b623.html.

[18] Schermann, M. Hemsen, H. Buchmüller, C. Bitter, T. Krcmar, H. Markl, V. and Hoeren, T. (2014). Big Data, An Interdisciplinary Opportunity for Info. Systems Research, DOI 10.1007/s12599-014-0345-1, Springer Fachmedien Wiesbaden.